

## Technical Note: A Blind Test of Mandibular Ramus Flexure as a Morphologic Indicator of Sexual Dimorphism in the Human Skeleton

STEVEN M. DONNELLY,<sup>1\*</sup> SAMANTHA M. HENS,<sup>2</sup>  
NIKKI L. ROGERS,<sup>2</sup> AND KENNAN L. SCHNEIDER<sup>2</sup>

<sup>1</sup>Department of Anthropology, University of Utah, Salt Lake City, Utah

<sup>2</sup>Department of Anthropology, University of Tennessee, Knoxville, Tennessee

**KEY WORDS** sex assessment; skeletal biology; blind test; interobserver reliability

**ABSTRACT** Loth and Henneberg (1996, *Am. J. Phys. Anthropol.* 99:473–487) identified a single morphological feature of the mandible, the presence or absence of a distinct flexure or angulation of the posterior margin of the mandibular ramus at the level of the occlusal plane, which appears to be an extraordinarily accurate predictor of sex. Using only this feature, Loth and Henneberg were able to predict sex with 94% accuracy in a large sample of mandibles. In this article, we report the results of a blind test of mandibular ramus flexure as a predictor of sex. In our blind test, only 62.5% of the mandibles were correctly sexed, and virtually identical results were obtained when the same sample of mandibles was examined by a second observer. Overall, our results demonstrate that: 1) the association between ramus flexure and sex is weak; 2) the predictive accuracy of Loth and Henneberg's method is better than chance for only one sex, males; and 3) the method is based on a trait that cannot be reliably or consistently identified. *Am J Phys Anthropol* 107:363–366, 1998 © 1998 Wiley-Liss, Inc.

Loth and Henneberg (1996) identified a morphologic feature of the mandibular ramus which appears to be an extraordinarily accurate predictor of sex. In examining a large sample of mandibles they found that adult males typically exhibit a distinct angulation of the posterior border of the vertical ramus at the level of the occlusal surface. In contrast, in the majority of females the posterior border of the ramus is straight, or if flexure is present it occurs at a higher point near the condyle, or lower in the mandible in association with gonial prominence or eversion. Using this feature alone, Loth and Henneberg were able to determine sex with 91 to 99% accuracy in skeletal samples representing African Blacks, Native Americans, American Whites, and American Blacks, with an overall accuracy of 94% for all samples combined. At 94%

accuracy, this method is as accurate as determining sex from the complete pelvis (St. Hoyme and İşcan, 1989).

Koski (1996) however, has questioned the reliability of ramus flexure as an indicator of sex, noting that very often it is difficult to distinguish its location and degree. In addition, the accuracy of the method has never been verified. In this article, we report the results of an independent, blind test of Loth and Henneberg's method.

### MATERIALS AND METHODS

To provide an independent assessment of ramus flexure as an indicator of sex, we conducted a blind test. The sample consists of 96 complete mandibles. No mandible

\*Correspondence to: Steven M. Donnelly, Department of Anthropology, 102 Stewart Building, University of Utah, Salt Lake City, UT 84112. E-mail: steven.donnelly@anthro.utah.edu

Received 29 October 1997; accepted 8 July 1998.

showed any obvious pathologies and none had more than one posterior tooth per side missing antemortem. Mandibles with post-mortem tooth loss were included in the sample only if at least one tooth remained on each side so that the level of the occlusal surface could be determined. Sixteen mandibles (two female and 14 male) from the University of Tennessee forensic collection and of known sex were used. This sample includes American Whites, American Blacks, and one Hispanic. The remaining 80 mandibles (38 female and 42 male) are of Native Americans. For these specimens, sex was determined from the morphology of the pelvis. The youngest individual in the forensic sample is 25 and all of the mandibles in the Native American sample are adults, as determined by eruption of the  $M_3$  and fusion of all long bone epiphyses.

The blind observer (Observer 1: KLS) was a student from a human osteology class who did not know of Loth and Henneberg's method, nor was she familiar with other methods of determining sex from the mandible. The observer was told that the purpose of the study was to determine whether flexure of the mandibular ramus could be used to predict sex. She was then shown mandibles exhibiting the different morphologies and these were also defined for her using the descriptions provided by Loth and Henneberg. She was not told which character states are believed to be associated with which sex. The observer was then asked to score the 96 mandibles using the same scoring system employed by Loth and Henneberg. That is, a score of +1 is assigned if the posterior margin of the ramus exhibits flexure at the occlusal level and a score of -1 is assigned if the posterior margin is straight, or if flexure is above or below the occlusal level. If the posterior margin of the ramus is neither clearly flexed nor straight a score of 0 is assigned.

The observer was presented the mandibles one at a time and asked to score one side. Only the ramus on the side being scored was visible, the remainder of the mandible being concealed by a sleeve. The mandibles were scored in four sessions on separate days, with left and right sides scored on different days. During these ses-

TABLE 1. Percentages of total (left + right) ramus flexure scores assigned by sex, and percentage accuracy of ramus flexure as a predictor of sex\*

| Score             | -2   | -1  | 0     | 1   | 2    | Accuracy |
|-------------------|------|-----|-------|-----|------|----------|
| <i>Observer 1</i> |      |     |       |     |      |          |
| Forensic sample   |      |     |       |     |      |          |
| Females           | 0.0  | 0.0 | 100.0 | 0.0 | 0.0  | 0.0      |
| Males             | 28.6 | 7.1 | 42.9  | 7.1 | 14.3 | 64.3     |
| Overall           |      |     |       |     |      | 56.3     |
| Amerindian sample |      |     |       |     |      |          |
| Females           | 34.2 | 5.3 | 31.5  | 5.3 | 23.7 | 39.5     |
| Males             | 14.3 | 0.0 | 38.1  | 0.0 | 47.6 | 85.7     |
| Overall           |      |     |       |     |      | 63.8     |
| Combined samples  |      |     |       |     |      |          |
| Females           | 32.5 | 5.0 | 35.0  | 5.0 | 22.5 | 37.5     |
| Males             | 17.9 | 1.8 | 39.3  | 1.8 | 39.3 | 80.4     |
| Overall           |      |     |       |     |      | 62.5     |
| <i>Observer 2</i> |      |     |       |     |      |          |
| Forensic sample   |      |     |       |     |      |          |
| Females           | 50.0 | 0.0 | 50.0  | 0.0 | 0.0  | 50.0     |
| Males             | 50.0 | 0.0 | 26.8  | 0.0 | 21.4 | 50.0     |
| Overall           |      |     |       |     |      | 50.0     |
| Amerindian        |      |     |       |     |      |          |
| Females           | 50.0 | 0.0 | 31.5  | 2.6 | 15.8 | 50.0     |
| Males             | 9.5  | 0.0 | 26.2  | 0.0 | 64.3 | 90.5     |
| Overall           |      |     |       |     |      | 71.3     |
| Combined          |      |     |       |     |      |          |
| Females           | 50.0 | 0.0 | 32.5  | 2.5 | 15.0 | 50.0     |
| Males             | 19.6 | 0.0 | 26.8  | 0.0 | 53.6 | 80.4     |
| Overall           |      |     |       |     |      | 67.7     |

\* Mandibles with total scores of -2 or -1 are identified as female, mandibles with total scores of 0, 1, or 2 are identified as male.

sions the scores from earlier sessions were not available to her. The first two sessions were for practice only. Total scores were calculated following Loth and Henneberg: for each mandible the scores for the left and right ramus are added and mandibles with scores of 0 through +2 are identified as males, and mandibles with scores of -1 or -2 are identified as females. In addition, the same sample was also scored by SMH (Observer 2), with left and right sides scored on separate days.

## RESULTS AND DISCUSSION

For our blind test, the prediction accuracy is much lower than that reported by Loth and Henneberg (Table 1). In presenting our results, we give the 95% confidence intervals in parentheses. In the blind test, only 62.5% (52.8–72.2%) of the mandibles were correctly sexed. For the males, the prediction accuracy was 80.4% (69.9–90.9%) but for the females it was only 37.5% (22.3–52.7%). Virtually the same results were obtained when the mandibles were scored by SMH. In this case 67.7% (58.3–77.1%) of the man-

TABLE 2. Joint proportions of scores for ramus flexure

|            | Side  | Observer 1 |      |      |       |      |      |
|------------|-------|------------|------|------|-------|------|------|
|            |       | Left       |      |      | Right |      |      |
|            | Score | -1         | 0    | 1    | -1    | 0    | 1    |
| Observer 2 | -1    | .258       | .000 | .144 | .320  | .000 | .155 |
|            | 0     | .021       | .000 | .000 | .041  | .000 | .010 |
|            | 1     | .134       | .000 | .442 | .165  | .010 | .299 |

dibles were correctly sexed. Sex was correctly predicted for 80.4% (69.9–90.9%) of the males, but the prediction accuracy for the females was only 50.0% (34.3–65.7%). Accuracy is consistently low for both the forensic sample, for which sex is known, and the Amerindian sample, for which sex is estimated from the pelvis. For the forensic sample, the accuracy is 56.3% (31.1–81.4%) for Observer 1, and 50% (24.7–75.3%) for Observer 2. For the Amerindian sample, the accuracy is 63.8% (53.1–74.4%) for Observer 1, and 71.3% (61.3–81.2%) for Observer 2. The low overall accuracy suggests that the association between ramus flexure and sex is weak, a conclusion that is supported by the results in Table 1, which show that a large proportion of the males have total scores between -2 and 0, and a large proportion of females have total scores between 0 and 2.

The joint proportions of scores assigned by the two observers are presented in Table 2. The relatively large values for many of the elements off the main diagonals indicate that the interobserver reliability of the method is low. The average value of kappa ( $\bar{\kappa}$ ), which is a measure of interobserver agreement (Fleiss, 1981), is 0.396 for the left side and only 0.280 for the right. Both values are statistically significant ( $P < 0.01$ ), indicating that agreement between the observers is greater than expected by chance. However, values of  $\bar{\kappa}$  less than 0.40 are generally considered to represent poor agreement beyond chance (Fleiss, 1981). Our finding of low interobserver agreement supports Koski's (1996) claim that the location and degree of ramus flexure is often difficult to identify.

In interpreting our results, it is necessary to take into account the fact that Loth and Henneberg's method is biased in favor of males. That is, individuals that exhibit the

putative female morphology are assigned a score of -2 or -1 and classified as females, individuals that exhibit the putative male morphology are assigned a score of +1 or +2 and identified as males, and individuals that do not clearly or consistently exhibit either morphological pattern on both sides are assigned a score of 0 and also classified as males. Given this scoring system, we could expect to achieve 60% accuracy for males by just randomly assigning scores to individuals. On the other hand, for females we should obtain only about 40% prediction accuracy by random chance. Furthermore, given the bias in favor of males in our sample and in the scoring system we would expect to achieve an overall prediction accuracy of 51.7% for our sample if scores were randomly assigned. Note that for the overall prediction accuracies that we obtained, the lower bounds of the confidence intervals are greater than the accuracy that could be expected by random chance. This is also the case for males. For females the 40% accuracy expected by random chance falls within the 95% confidence intervals for the actual prediction accuracies obtained by both observers.

The poor predictive ability of Loth and Henneberg's method is not the result of flaws in the scoring system, and altering the system does not improve the overall accuracy of the method. If mandibles with total scores of -2 through 0 are classified as females and mandibles with scores of 1 or 2 are identified as males the overall accuracy is 54.2% (44.1–64.2%) for Observer 1, and 65.6% (56.1–75.2%) for Observer 2. Since the scoring system is now biased in favor of females, the predictive accuracy for females increases to 72.5% (58.4–86.5%) for Observer 1, and 82.5% (70.6–94.4%) for Observer 2, but for males it drops to 54.2% (44.1–64.2%) for Observer 1, and 53.6% (40.4–66.7%) for Observer 2.

## CONCLUSION

Using only ramus flexure as an indicator of sex, and following the methods outlined by Loth and Henneberg (1996), we obtained a prediction accuracy of only 63–69%. This is much lower than the 91–99% accuracy reported by Loth and Henneberg. It is also

considerably less than the 85% accuracy of discriminant functions for determining sex from the mandible (Giles, 1964). Our results do show that Loth and Henneberg's method can be used to predict sex better than the approximately 52% accuracy that we would expect to achieve for our sample if scores had been randomly assigned. Our results also show that using Loth and Henneberg's scoring system, which is biased in favor of males, there is an approximately 80% chance that a male mandible will be correctly identified as such. This is significantly better than the 60% accuracy that could be obtained by random chance. However, there is only a 50% or less chance of correctly identifying a female mandible as such, and the prediction accuracies that we obtained for females are not significantly different from the 40% accuracy that would be expected if scores were assigned randomly. In addition, interobserver agreement for the scores is low, indi-

cating that the presence and location of ramus flexure is often difficult to identify (cf., Koski, 1996). From a practical standpoint, a method of determining sex that is based on a trait that cannot be reliably or consistently identified, that has an overall accuracy of less than 70%, and whose predictive ability is better than chance for only one sex is of questionable utility.

#### LITERATURE CITED

- Fleiss JL (1981) *Statistical Methods for Rates and Proportions*. New York: John Wiley & Sons.
- Giles E (1964) Sex determination by discriminant function analysis of the mandible. *Am. J. Phys. Anthropol.* 22:129–136.
- Koski K (1996) Mandibular ramus flexure—Indicator of dimorphism? *Am. J. Phys. Anthropol.* 101:545–546.
- Loth SR, and Henneberg M (1996) Mandibular ramus flexure: A new morphologic indicator of sexual dimorphism in the human skeleton. *Am. J. Phys. Anthropol.* 99:473–487.
- St. Hoyme L, and Isçan MY (1989) Determination of sex and race: Accuracy and assumptions. In MY Isçan and KAR Kennedy (eds.): *Reconstruction of Life from the Skeleton*. New York: Alan R. Liss, pp. 53–94.